A. Introduction Section

The National Health and Nutritional Examination Survey provided information for this data set. The information was collected on adults who are at an age of 20 years or older between the years of 1988 and 1994.

The goal of this data set is to provide the Blood Pressure Rank, measured by variable SBPRANK, for each subject in the Data Set. SBPRank is the grouping variable that is determined by average systolic blood pressure (AVGSBP). The values for SBPRank include 1,2, and 3. 1 is the lowest blood pressure rank indicated by low values of AVGSBP and 3 is the highest blood pressure rank indicated by high values of AVGSBP. 2 is the blood pressure ranking for average systolic blood pressure in the middle ranges.

The other variables in the data set includes:

- HSAGEIR: Subject age recorded in years.
- BMPWTLBS: Subject body weight in pounds.
- BMPHTIN: Subject standing height in inches.
- PEPMNK5R: Subject average diastolic blood pressure
- TCP: This is a measure of the Subject Serum Cholesterol.

C1: Correlated Quantitative Variables (Multicollinearity)

The Correlation Coefficient Tables for between quantitative variables in the NHANES data set are below:

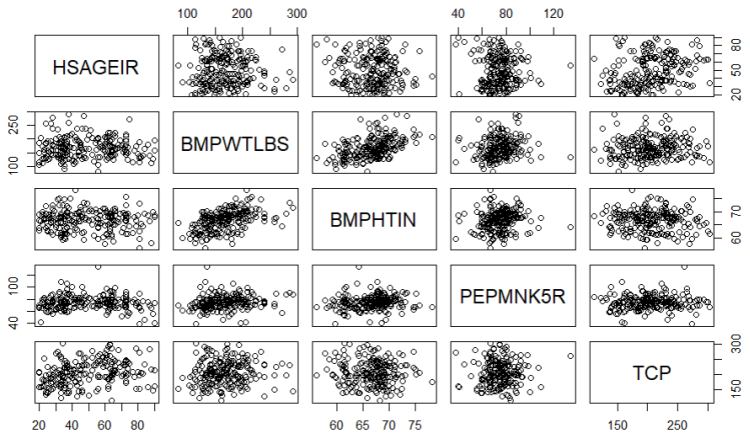
Subsetting the Data instead of doing column removal to get correlation matrix:

```{r}
#obtaining the correlation Matrix
cor(NHANESDataSet[,c("HSAGEIR", "BMPWTLBS", "BMPHTIN", "PEPMNK5R", "TCP")])
```

	HSAGEIR	BMPWTLBS	BMPHTIN	PEPMNK5R	TCP
HSAGEIR	1.000000000	-0.008383008	-0.06742182	-0.0729738	0.28193987
BMPWTLBS	-0.008383008	1.000000000	0.48165357	0.1720942	0.08077624
BMPHTIN	-0.067421821	0.481653572	1.00000000	0.1387325	-0.06180148
PEPMNK5R	-0.072973801	0.172094241	0.13873253	1.0000000	0.04779740
TCP	0.281939870	0.080776245	-0.06180148	0.0477974	1.00000000

The Scatter Plots representing correlations between the quantitative variables in the NHANES data set are below:

```{r}
plot(NHANESDataSet[,c("HSAGEIR", "BMPWTLBS", "BMPHTIN", "PEPMNK5R", "TCP")])
```



R Code Commands:

Reading in the NHANES data set

```r
{r}
NHANESDataSet <- read.csv("C:\\Users\\tarar\\Downloads\\NHANES3_419.csv")
NHANESDataSet
```

Description: df [209 × 6]

SBPRANK <int>	HSAGEIR <int>	BMPWTLBS <dbl>	BMPHTIN <dbl>	PEPMNK5R <int>	TCP <int>
1	63	141.4	63.5	64	220
1	28	142.2	61.2	77	218
1	25	136.5	64.4	59	138
1	33	127.7	68.1	65	252
1	47	134.8	59.2	79	200
1	55	159.5	67.8	65	240
1	32	128.8	63.7	66	188
1	74	123.0	61.5	53	241
1	20	113.5	62.8	67	159
1	43	198.0	63.1	79	164

#obtaining the correlation matrix

```
cor(NHANESDataSet[,c("HSAGEIR", "BMPWTLBS", "BMPHTIN", "PEPMNK5R", "TCP")])
```

#making the correlation plot with code below:

```
plot(NHANESDataSet[,c("HSAGEIR", "BMPWTLBS", "BMPHTIN", "PEPMNK5R", "TCP")])
```